Jurnal Psikologi
Volume 52, Number 1, 2025: 1– 14
DOI: 10.22146/jpsi.99403

ISSN 0215-8884 (print)
ISSN 2460 867X (Online)

Psychometric Properties of the Trail-Making Test in Dementia Population

Rahmadi Maulana Kusuma Wahyudi*[1], Evany Victoriana[1], C.M. Indah Soca R. Kuntari[1], Paulus Anam Ong[2], Augustina Sulastri[3], Gilles van Luijtelaar[1,4]

[1]*Faculty of Psychology, Maranatha Christian University, Indonesia*

[2]*Department of Neurology, Hasan Sadikin Hospital, Indonesia*

[3]*Faculty of Psychology, Soegijapranata Catholic University, Indonesia*

[4]*Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Netherlands*

Submission 25 August 2024 **Accepted 17 February 2025** **Published 30 April 2025**

Abstract. The trail-making test (TMT) is a commonly used executive function test in various populations worldwide. It is also used to assess the dementia population, who experience impairment in executive function. In Indonesia, the psychometric properties of TMT for healthy subjects local populations have been reported to be satisfactory. However, there is a lack of studies on the validity and reliability of the TMT for the dementia population. Here we examined the validity and reliability of TMT and established population-specific norms for dementia. We analyzed the TMT scores and three other neuropsychological tests (DSST, DS test, and MoCA) from 71 subjects with a clinical diagnosis of dementia scores and three other neuropsychological tests (DSST, DS test, and MoCA) and examined its psychometric properties. The internal consistency analysis was used to assess the reliability of the TMT, while a generalized structured component analysis (GSCA) was conducted to examine TMT's validity in assessing executive function. The reliability test indicated that the TMT has a high reliability (ω = 0.639), while GSCA revealed that TMT (TMT-A=0.72 and TMT-B=0.573) has high construct validity with three other neuropsychological tests. Additionally, this study provided standardized norms of TMT for the dementia population.

Keywords: dementia; executive functions; psychometric properties; trail making test

Neuropsychological assessments are used to evaluate cognitive abilities. These tests are highly practical, allowing professionals to identify irregularities and describe the severity patterns. Additionally, neuropsychological test plays an important role in establishing diagnoses and explaining cognitive effects associated with known neuropsychological conditions (Elkana et al., 2015; Zucchella et al., 2018). Today, there is a growing focus on research about neuropsychological tests. In recent years, several universities in Indonesia have collaborated to form a neuropsychology consortium, leading to the development of the Indonesian Neuropsychological Test Battery (INTB), which consists of widely

*Address for correspondence: 2232024@psy.maranatha.edu

used instruments. This test battery comprises ten test kits designed to measure cognitive abilities in specific brain domains, namely the Boston Naming Test, Digit Span, RAVLT, Five Point Test, phonemic Verbal Fluency Test, Stroop Colour Word Test, Figural Reproduction Test, Bourdon Wiersma Test, Token Test, and Trail Making Test (Wahyuningrum et al., 2022).

The Trail Making Test (TMT) is a neuropsychological measurement designed to measure various domains of executive function, e.g., psychomotor speed, visual attention, cognitive flexibility, processing speed, sequencing, mental flexibility, and visual-motor skills. The test is used both in the healthy and clinical populations. The test comprises two parts, TMT-A and TMT-B and is relatively straightforward to comprehend and administer. In TMT-A, participants are instructed to connect numbers sequentially (1-2-3, and so forth) using uninterrupted lines at semi-random places on the test sheet. In TMT-B, participants follow the same instructions but with a combination of numbers and letters (1-A-2-B-3-C, and so forth) (Bezdicek et al., 2012; Mališová et al., 2021; Ojeda et al., 2014; Reitan & Wolfson, 2004; Seo et al., 2006; Siciliano et al., 2019; Specka et al., 2022; Wagner et al., 2011; Widhianingtanti et al., 2022).

In the neuropsychological testing discourse, psychometric properties of the test are essential. Analysis and evaluation of the validity, reliability, and standardized norm of an instrument are vital once the instrument is completed or adapted to ensure its quality (DeVellis & Thorpe, 2022; Souza et al., 2017). The psychometric properties of the TMT are reasonably well-documented for the healthy population in Indonesia. Number of studies have consistently reported positive results on clinical population, such as amnestic mild cognitive impairment (aMCI) and major depressive disorder population internationally (Bezdicek et al., 2012; Bracken et al., 2019; Mališová et al., 2021; Osuka et al., 2020; Specka et al., 2022; Wagner et al., 2011; Widhianingtanti et al., 2022). However, there is a lack of data regarding the psychometric properties of the TMT adjusted for a clinical population with cognitive impairment in Indonesia. It is important to understand that there are differences between neurotypical individuals and those with certain cognitive disorders. In this study, it was assumed that neurotypical individuals would complete the TMT quicker than individuals with cognitive impairments. This discrepancy results in difficulty in fairly evaluating individuals with cognitive disorders and their neurotypical counterparts. Therefore, it is necessary to differentiate between "normal" scores for the neurotypical population and "normal" scores for those with cognitive impairments. Considering assessment and intervention purposes for the clinical population, the psychometric properties of the TMT that are available for healthy populations may be unsuitable in clinical settings, therefore, an adaptation of this instrument for the clinical populations is required (Suwartono et al., 2014; Wahyuningrum et al., 2022). This population gap can be addressed through categorization tailored specifically to clinical populations.

Clinical populations with cognitive impairment are diverse in terms of irregularities in specific cognitive domains. Literature has suggested that dementia is becoming increasingly prevalent worldwide among clinical subjects with cognitive impairment, specifically among those with impairment in executive functions (Prince, 2015; Turana et al., 2019). Dementia serves as an umbrella term for various symptoms associated with impairments typically observed in older adults,

e.g., Alzheimer's disease, Parkinson's disease, as well as vascular and frontotemporal dementia. Individuals with dementia commonly see accelerated age-related declines in memory function and cognitive capabilities, with a varying age of onset. This condition is marked by various issues affecting behavior, memory, and cognitive abilities, resulting in a diminished capacity to perform daily activities. Dementia has an impact on multiple brain functions, including the executive function. Examples of impaired executive function in individuals with dementia encompass deficiencies in planning and decision-making, diminished working memory function and selective attention, weakened attention to details, failure to ignore distracting stimuli, and compromised manipulation skills (Ashraf & Uddin, 2022; Breijyeh & Karaman, 2020; Kirova et al., 2015; Mendez, 2017; Trull & Prinstein, 2013). A study by Alzheimer's Disease International (ADI) in 2015 estimated that 1.2 million people in Indonesia should have been diagnosed with dementia and this number may rise to 4 million by 2050 (Prince, 2015; Turana et al., 2019). There is an urgent need for early assessments to enable timely interventions for people with high risks of dementia before the disorder advances (Suriastini et al., 2018; Sweetasari, 2022).

Identifying the dementia population in healthcare facilities is relatively straightforward. In Bandung, Indonesia, Hasan Sadikin Hospital has a neurology department that specifically handles patients with dementia. Staff at this hospital employ diverse assessment techniques, including neuropsychological tests, e.g., the trail-making test (TMT), digit symbol substitution test (DSST), digit span (DS) test, and the Indonesian version of the Montreal cognitive assessment (MoCA). These tests are commonly used in clinical settings for brief assessment. In this study we used the available neuropsychological test data from the mentioned neurology department. There has been a lack of routine monitoring of dementia populations in Indonesia aside from Alzheimer's Disease International's studies, which makes it difficult to obtain a large sample for this type of study. Furthermore, the specific location of this study makes generalization limited to the dementia population at Hasan Sadikin Hospital.

Motivated by the findings and limitations mentioned above, we studied the psychometric properties of TMT in a dementia population in Indonesia. This initiative aimed to address gaps in previous studies' populations and offer preliminary data on adjusted norms. According to our literature review, there was no data available for the TMT for Indonesia's dementia population. 'Empirically approved instruments' with high-quality psychometric properties can assist clinical practitioners in conducting accurate assessments and implementing interventions at the earliest stages, which will help slow the progression of the disease.

Methods

This study adopted a quantitative research approach, employing a non-experimental design. All data used in this research were secondary data, provided by the neurology department of Hasan Sadikin Hospital through patient medical records from 2017 to 2021. The dataset included scores from the TMT Part A and B, DSST, DS Forward, and MoCA. Furthermore, demographic information, e.g., gender, age,

education level, and diagnosis, were integrated into the dataset.

Participants

The research was conducted at a hospital to specifically target individuals with dementia. Data from 71 patients from 2017 to 2021 were obtained from the medical records of the Hasan Sadikin Hospital, Bandung. Of the patients, 54.4% were female, aged from 43 to 89 years (M=67.4). Based on the data, we could conclude that the age range of the patients was relatively broad, spanning from middle to late adulthood. Participant ages were categorized as follows: 43–49 years, 50–59 years, 60–69 years, 70–79 years, and 80–85 years. The education years of the participants varied, ranging from 6 to 21 years (M=13.1, SD=3.07). Education data were not categorized into specific groups. Most patients in the study were diagnosed with Alzheimer's dementia (48.5%). The study adhered to the ethical guidelines of the Research Ethics Committee at Hasan Sadikin Hospital (research clearance number: DP.04.03/D.XIV.2.2.1/22583/2023, ethical approval number: LB.02.01/X.6.5/416/2023). Table 1 provides a demographic overview of the participants.

Table 1

Demographic Data of the Normative Group (N=71)

Variables	Category	*N*
Sex	Male	34
	Female	37
Age	43-49	4
	50-59	13
	60-69	19
	70-79	28
	80-85	7
Education (in years)	6	4
	9	5
	10	2
	12	9
	15	12
	16	14
	18	4
	21	1
Diagnosis	Alzheimer's Dementia	34
	Frontotemporal Dementia	1
	Parkinson's Dementia	5
	Vascular Dementia	31

Measures

The assessment was conducted by a medical doctor affiliated with the neurology department at Hasan Sadikin Hospital. Each participant underwent an evaluation using both segments of the TMT-A and B. Before the actual test, all participants had the opportunity to practice for both components. In

TMT-A, participants were instructed to sequentially connect numbers (1-25) within circles without making errors. For TMT-B, participants were instructed to connect a combination of numbers and letters (1 to 13 and A to L) within circles without errors as quickly as possible. If an error is made, the tester corrected it, tallied the total errors, and proceeded with the test without interrupting the time elapsed. The time limits for Part A and Part B were set at 180 seconds and 300 seconds, respectively. The other neuropsychological tests used in this study were DSST, DS Forward, and MoCA. Data scores of these tests were used for validity test, because the DSST also measures executive functions like the TMT, while the DS Forward measures an element of working memory, often considered as belonging to the executive function domain (Wulanyani et al., 2024). The MoCA on the other hand, consist of six subscales measuring various neuropsychological functions (attention via e.g. DS Forward, executive function e.g. via a short version of TMT B , language, visuospatial functions, memory, and orientation). The total score of the subtests of the MoCA was used, as commonly done. These other neuropsychological tests were also administered at the same time by the neurology department staff at Hasan Sadikin Hospital.

Statistical Analysis

The researchers calculated mean, median, standard deviation, skewness, kurtosis, maximum, minimum, and percentiles (25^{th}, 50^{th}, 75^{th}) for the entire sample to obtain raw normative data (Table 2). As there was no retest or additional data available, internal consistency analysis was employed to assess the reliability of the instrument, with cut-off scores 0.6 – 0.8 for high reliability according to Guilford's classification for reliability (Divayana et al., 2019; Price, 2017). The analysis was conducted using Jamovi software version 2.3.24.

Considering the small sample size and the non-normal data distribution, generalized structured component analysis (GSCA) was performed on the scores of TMT and other neuropsychological tests (DSST for convergent validity, DS Forward, and MoCA for discriminant validity) to establish the TMT's validity as executive function test. The analysis was done using the GSCA Pro software version 1.2.1.0. GSCA produces optimal estimation results because of its capability to analyze small data samples with no prerequisite test (Hwang, 2009). All data were standardized into Z-scores to facilitate accurate comparison and ensure equivalent interpretation. Lastly, the raw scores of the three variables of the TMT were analyzed using Jamovi's additional module: cNorms, to generate standardized norms for a single group for each variable (TMT-A, TMT-B, and TMT B-A).

Results

The researchers analyzed a total of 71 samples. The obtained data consisted of test scores of TMT-A, TMT-B, TMT B-A (calculated by subtracting TMT-B with TMT-A), DS Forward, DSST, and MoCA. Demographic data provides information on gender, age, education, and dementia diagnosis. The TMT data showed no missing scores in the dataset. The mean scores for TMT-A=103 (*SD*=45.3), TMT-B=252 (*SD*=65.0), and TMT B-A=149 (*SD*=58.5), the medium to large SD's indicate variability within the

dataset. The skewness and kurtosis of TMT-A indicated a right-skewed distribution (0.470) and a relatively flat peak (-0.893), while TMT-B showed a left-skewed distribution (-1.04) and a flatter kurtosis (-0.334). For the TMT B-A scores, data distribution was left-skewed (-0.0847) and showed a flatter kurtosis (-0.345), suggesting that the obtained data did not follow the normal distribution. Among the 71 data points collected, three samples had incomplete data (lacked DSST scores). Considering that individuals with dementia disorders might exhibit high scores on the tests, no data were excluded from the analysis. Table 2 presents score descriptions for the three TMT variables, DS Forward, DSST, and MoCA (*N*=71).

Table 2

Descriptive Statistics of TMT, DS Forward, DSST, and MoCA (N=71)

	TMT-A	TMT-B	TMT B-A	DS Forward	DSST	MoCA
N	71	71	71	71	68	71
Missing	0	0	0	0	3	0
Mean	103	252	149	4.25	22.3	15.1
Median	91	300	141	4	22.0	16
SD	45.3	65.0	58.5	1.02	8.79	3.70
Min.	12	85	6	2	0	3
Max.	180	300	288	6	43	25
Skewness	0.470	-1.04	-0.0847	-0.370	-0.0639	-0.496
Std. error skewness	0.285	0.285	0.285	0.285	0.291	0.285
Kurtosis	-0.893	-0.334	-0.345	-0.350	-0.285	1.01
Std. error kurtosis	0.563	0.563	0.563	0.563	0.574	0.563
25^{th} percentile	70.0	211	117	4.00	15.0	13.0
50^{th} percentile	91.0	300	141	4.00	22.0	16.0
75^{th} percentile	133	300	195	5.00	29.0	18.0

Reliability

Table 3 presents the results of the internal consistency analysis, which revealed the instrument's reliability value. The TMT (Part A and B) was considered to have high reliability (ω = 0.653).

Table 3

Internal Consistency of TMT-A and TMT-B (N=71)

	Mean	*SD*	McDonald's ω
Scale	178	47.7	0.653

Validity

The construct validity analysis was conducted using the GSCA technique, specifically the basic GSCA/single group analysis on time scores of the TMT and three other neuropsychological tests: DS forward (assessing attention and working memory), DSST (measuring processing speed, working memory, visuospatial processing, attention), and MoCA (memory, executive function, attention, language, visuo-spatial functions, and orientation). The data were also transformed into Z-scores for

standardization and to facilitate meaningful comparisons. GSCA does not require the assumption of data normality so the researchers did not need to perform the normality test. The primary results of the analysis are presented in Table 4.

Table 4

Results of Component Analysis of Five Neuropsychological Tests (N=71)

	Estimate		SE		95% CI			
ONT	ω	c	ω	c	ω	ω	c	c
DS Forward	0.226	0.339	0.161	0.209	-0.131	0.478	-0.134	0.658
DSST	0.751	0.848	0.094	0.091	0.5	0.916	0.606	0.956
MoCA	0.47	0.61	0.103	0.119	0.229	0.633	0.324	0.769
TMT								
TMT Time A	0.72	0.826	0.088	0.067	0.56	0.93	0.67	0.95
TMT Time B	0.573	0.706	0.101	0.112	0.312	0.748	0.378	0.032

Note: ONT = other neuropsychological tests, ω = weights, c = loadings. Data were already standardized using Z-score.

Based on Hwang et al. (2023) recommendation for a small sample size, the model obtained from GSCA had a good fit (GFI=0.975, SRMR=0.072). The validity analysis also found that TMT-A (ω= 0.72, c= 0.826) and TMT-B (ω= 0.573, c= 0.706) had satisfying validity. The TMT B-A data were excluded in the GSCA analyses considering that its inclusion resulted in multicollinearity.

The result demonstrated that the TMT-A (ω= 0.72, c= 0.826) and TMT-B (ω= 0.573, c= 0.706) has convergent validity with the DSST (ω= 0.751, c= 0.848) and MoCA (ω= 0.47, c= 0.61), while DS Forward (of ω= 0.226, c= 0.339) showed discriminant validity toward TMT. However, among the tests, DS Forward was not found to have significant weights and loadings (95% Cl < 0). This finding indicates that TMT, DSST, and MoCA measure rather similar elements of executive functions, while the low weights and loadings between TMT and DS Forward indicate that TMT measures a different construct as the DS Forward. This confirmed the validity of TMT in assessing a patient's neuropsychological condition.

Standardized Norm

The minimum, 25th percentile, median, 75th percentile, and maximum scores of the three TMT variables (Table 2) were used to establish raw norms or raw scores for the TMT. These values were arranged into five categories: very high, high, moderate, low, and very low. A higher score reflected a higher severity of dementia, and vice versa. The raw scores were then processed using a norm score generator for single groups with the cNorm module of the Jamovi's software. The output of this process was a standardized norm in the form of T-scores. The distribution of the scores followed the raw norm categories in a reverse order, meaning the higher the raw norm score, the lower the standardized norm score, and vice versa. Table 5 presents categorized raw norms and standardized norms.

Table 5

Categories of TMT-A, TMT-B, and TMT B-A Scores

Category	TMT-A		TMT-B		TMT B-A	
	Raw score	T-score	Raw score	T-score	Raw score	T-score
Very High	>180	-	-	-	>288	-
High	134 – 180	43.8 – 35.8	>300	-	196 – 288	42.3 — 21.7
Moderate	92 – 133	50.2 – 44.0	212 – 300	56.3 – 49.4	142 – 195	51.2 – 42.5
Low	71 – 91	55.2 – 50.3	85 – 211	74.5 – 56.3	118 – 141	55.1 – 51.4
Very Low	12 – 70	74.2 – 55.1	-	-	6 – 117	74.4 – 55.3

Discussion

The lack of psychometric properties data for neuropsychological tests among the healthy population in Indonesia has raised concerns among psychologists. This is particularly challenging due to the difficulty of obtaining similar normative data for clinical populations. While comparable studies had been conducted internationally, to the best of the researcher's knowledge, this marks the first instance of cognitive assessment within the dementia population in Indonesia. It should be noted that the number of included subjects was not large, therefore generalizability of the outcomes of this research might be limited and only valid for the dementia populations in Hasan Sadikin Hospital. On the other hand, dementia is a global phenomenon and there are no indications that the population in the Hasan Sadikin Hospital is atypical. Furthermore, the dataset encompassed dementia patients across four years, meaning that the dataset had more spread, particularly the TMT data ($SD_{TMT\text{-}A}$= 45.3, $SD_{TMT\text{-}B}$= 65.0, $SD_{TMT\ B\text{-}A}$= 58.5, $SD_{DS\ Forward}$= 1.02, SD_{DSST}= 8.79, SD_{MoCA}= 3.70).This indicates high variability within the dataset and, therefore, the results must be confirmed in future studies using larger, homogeny and normally distributed samples of the dementia population.

A comparative analysis of the mean and percentile data from this study's participants on the three variables of TMT (A, B, B-A) revealed significant deviations from normative performance as reported in previous literature. The findings suggest markedly reduced cognitive functioning among participants in this sample. The mean completion time for all three TMT variables in the current study (M=103, 252, 149) was substantially longer than the normative means reported by Wahyuningrum et al. (2022) and Widhianingtanti et al. (2022). They reported cut-off scores of 81, 191 and 114 seconds respectively for the three TMT variables based on a Javanese sample. An analysis of the percentile distributions further underscores the severity of cognitive slowing in the current sample, indicating that a significant portion of participants could not complete the task within the limits set by the norm group. A more detailed comparison with the normative data is less meaningful considering the lack of age adjusted norms scores for the elder healthy population. These findings demonstrate that a large part of distribution of scores in this "clinical" samples falls outside the "normal" range established by prior studies. The consistent deviation across both mean and percentile metrics suggests an obvious impairment in dementia population compared to healthy population.

The reliability test results showed that the TMT had a high reliability (ω = 0.639) when used

in the dementia population. The findings of this study were similar to its predecessors, which studied healthy and other clinical populations (Widhianingtanti et al., 2022) reported that TMT-A (r_s= 0.76), TMT-B (r_s=0.86), and TMT B-A (r_s=0.74) for the healthy population in Indonesia had satisfying intraclass correlation, showcasing their reliability. A study on patients with major depressive disorder in Germany also found that TMT-A and TMT-B had satisfying reliability (Wagner et al., 2011). Another study in Japan (Osuka et al., 2020) on older adults found similar results. In the United States, Bracken et al. (2019) on healthy participants also found satisfying reliability for both TMT-A and TMT-B. Based on the literature review, a commonality among previous studies is the absence of internal consistency testing. Most previous studies employed techniques, such as test-retest, alternate form, and intraclass correlation coefficient (ICC). The findings of this research contribute empirical evidence to enhance the reliability metrics of several established reliability testing methods. Moreover, the study confirmed that the TMT maintains a high level of reliability.

The validity test results, obtained through GSCA showed that TMT-A (ω=0.72, c=0.826) and TMT-B (ω=0.573, c=0.706) could effectively asses one of the three recognized component in executive function, i.e. cognitive flexibility, rather than working memory or the inhibitory control component in the dementia population (Wulanyani et al., 2024). TMT-A and TMT-B also exhibited a robust association with the DSST (ω=0.751, c=0.848). Meanwhile, no significant association was observed with the DS Forward. The analysis results confirmed the TMT's construct validity, along with its convergent and discriminant validity.

Both TMT parts assess psychomotor speed, visual attention, executive function, cognitive flexibility, processing speed, sequencing, mental flexibility, and visual-motor skills. DSST measures motor speed, attention, visuoperceptual functions, scanning, writing/drawing, planning, strategizing, and working memory. The findings of this study underscore that despite measuring different constructs, both tests can effectively assess executive function, indicating convergent validity. The results align with Widhianingtanti et al. (2022), confirming that the executive function domain encompasses multiple cognitive abilities processed simultaneously, without differentiation into distinct constructs. Primarily the weights of DS Forward and MoCA (ω<0.5) indicate no significant relationship with TMT within the model. This signifies discriminant validity, suggesting that TMT discriminates and does not measure variables evaluated by DS Forward and MoCA.

However, this contradicts previous research findings about a strong association between TMT and DS test (Widhianingtanti et al., 2022). To address this contradiction, it is important to comprehend the three forms of DS tests: Forward, Backward, and Sequence. The present study used the DS Forward score, while Widhianingtanti et al. (2022) employed DS Backward and Sequence. These distinct scores involve different cognitive processes. The researchers posited that the score difference resulted in the insignificant DS Forward. DS Forward is considered more sensitive to the short-term memory domain and attention, which aligns with the measured constructs (attention and working memory). Meanwhile, DS Backward and Sequence are presumed to require higher cognitive processes, contributing to the executive function domain, including sequencing and mental flexibility, which are less necessary in performing well in a short-term memory test. The MoCA, on the other hand, being

a concise test measuring different domains, includes the alternating trail-making subtest (similar to TMT-B, but much shorter), the attention subtest covering DS Forward and Backward and other subtests that measure other cognitive domains, leading to a low loading factor output. (Carson et al., 2018; Nasreddine et al., 2005).

Moreover, the method employed in this study differs from previous research. This study involved a population with more specific disorders, a smaller sample size, and a more flexible analytical approach, which likely contributed to the divergent results. The contradictory findings open new opportunities and research areas to further explore the validity of the DS test in measuring learning and memory or attention. Additionally, it suggests the potential for the DS test to assess executive function domains, similar to the TMT. The MoCA instrument should be examined further, such as by analyzing not just the total score, but rather the individual subtest scores in a more granular manner, comparing them to other assessment tools based on the specific domains measured by each MoCA's subtest.

Despite previous findings, overall, this study demonstrated good validity and reliability of TMT, allowing the researchers to establish standardized norms for the dementia population, although with limited applicability to the Hasan Sadikin Hospital setting. The introduction of standardized norms provides a crucial framework, particularly for clinical practitioners, by establishing a standard or baseline for more objective, reliable, and accurate evaluations of dementia patients. This finding represents an important step toward implementing both preventive and curative interventions for the dementia population. Given the rising prevalence of dementia globally, especially in Indonesia, it is essential to conduct comprehensive assessments of the severity of the disease in patients. Such assessments enable clinical practitioners to deliver more appropriate treatments based on the severity levels. Furthermore, the development of standardized norms for the TMT could be further extended by considering additional demographic factors, e.g., gender, age, types of dementia, cultural differences, and educational level. In this way, the TMT—although a relatively simple assessment tool—could evolve into a more holistic instrument for evaluating executive functions within the dementia population.

Conclusion

This study offers important insights into the reliability and validity of the TMT for assessing executive function in Indonesia's dementia population. The results confirmed the measurement's high reliability, consistent with previous studies in healthy and clinical populations. It also confirmed the measurement's strong construct validity, supporting its role in measuring various aspects of executive function, e.g., cognitive flexibility and sequencing.

However, the difference with previous research regarding the relationship between TMT and DS Forward scores highlight methodological differences, suggesting that while DS Forward measures attention and memory, other forms of the DS test (backward, sequence) might better assess executive function. Additionally, the study's moderate sample size and limited applicability to the dementia

population at Hasan Sadikin Hospital suggest that future research with larger, more homogeneous samples is needed to confirm these findings. The establishment of standardized norms for the TMT within this clinical context represents a crucial step in providing clinicians with a reliable baseline for evaluating dementia patients. This will support more accurate, objective assessments and personalized treatment plans, which are increasingly important given the rising prevalence of dementia in Indonesia. Future research should further explore the impact of demographic factors (e.g., age, gender, dementia type) on TMT performance, expanding its use as a more comprehensive tool for assessing executive function in diverse populations. In conclusion, this study lays the groundwork for advancing neuropsychological testing in dementia care, offering practical insights into improving diagnostic and therapeutic practices.

Recommendation

Future research should involve larger, more diverse samples to improve the generalizability of findings and establish more robust normative data for dementia populations across Indonesia. It is recommended to further validate the TMT in other clinical populations and settings to expand its use as a reliable measure of executive functions in dementia patients. Furthermore, development of the standardized norms for the TMT that consider age, gender, severity and dementia type, culture, and education level should enhance its applicability across diverse dementia populations. In clinical practice, practitioners should adopt standardized TMT norms to improve diagnostic accuracy and treatment planning, supporting individualized care for dementia patients.

Declaration

Acknowledgments

Authors express gratitude for the contributions to Maranatha Christian University, Arief Budiarto, Efi Fitriana, and Tery Setiawan.

Funding

The research was funded by Maranatha Christian University.

Authors' Contributions

RMK conceived and designed the study, wrote the manuscript, organized the data, and analyzed the data. EV conceived and designed the study, supervised and reviewed the mauscript draft. CMISRK conceived and designed the study, and supervised and reviewed the manuscript draft. PAO conceived the study of dementia, contributed to method planning, and organized the data collection. AS conceived the Trail Making Test study. GVL conceived the study, supervised and reviewed the mauscript draft.

Conflict of Interest

The authors declare not to have any competing interests related to this work.

Orcid ID

Rahmadi Maulana Kusuma Wahyudi https://orcid.org/0009-0000-9657-8234
Evany Victoriana https://orcid.org/0009-0004-9493-5436
C.M. Indah Soca R. Kuntari https://orcid.org/0000-0001-9658-0852
Paulus Anam Ong https://orcid.org/0000-0002-0787-2907
Augustina Sulastri https://orcid.org/0000-0002-0107-7590
Gilles van Luijtelaar https://orcid.org/0000-0002-0710-3403

References

Ashraf, G. M., & Uddin, M. S. (2022). *Current thoughts on dementia: From risk factors to therapeutic interventions*. Springer Nature Singapore. https://doi.org/10.1007/978-981-16-7606-2

Bezdicek, O., Motak, L., Axelrod, B. N., Preiss, M., Nikolai, T., Vyhnalek, M., Poreh, A., & Ruzicka, E. (2012). Czech version of the trail making test: Normative data and clinical utility. *Archives of Clinical Neuropsychology*, *27*(8), 906–914. https://doi.org/10.1093/arclin/acs084

Bracken, M. R., Mazur-Mosiewicz, A., & Glazek, K. (2019). Trail making test: Comparison of paper-and-pencil and electronic versions. *Applied Neuropsychology: Adult*, *26*(6), 522–532. https://doi.org/10.1080/23279095.2018.1460371

Breijyeh, Z., & Karaman, R. (2020). Comprehensive review on Alzheimer'S Disease. *World Journal of Pharmacy and Pharmaceutical Sciences*, *10*(7), 1170. https://doi.org/10.20959/wjpps20217-19427

Carson, N., Leach, L., & Murphy, K. J. (2018). A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *International Journal of Geriatric Psychiatry*, *33*(2), 379–388. https://doi.org/10.1002/gps.4756

DeVellis, R. F., & Thorpe, C. T. (2022). *Scale development: Theory and applications*. SAGE Publications, Inc.

Divayana, D. G. H., Adiarta, A., & Gede Sudirtha, I. (2019). Instruments development of tri kaya Parisudha-based countenance model in evaluating the blended learning. *International Journal of Engineering Pedagogy*, *9*(5), 55–74. https://doi.org/10.3991/ijep.v9i5.11055

Elkana, O., Eisikovits, O. R., Oren, N., Betzale, V., Giladi, N., & Ash, E. L. (2015). Sensitivity of neuropsychological tests to identify cognitive decline in highly educated elderly individuals: 12 months follow up. *Journal of Alzheimer's Disease*, *49*(3), 607–616. https://doi.org/10.3233/JAD-150562

Hwang, H. (2009). Regularized generalized structured component analysis. *Psychometrika*, *74*(3), 517–530. https://doi.org/10.1007/S11336-009-9119-Y

Hwang, H., Cho, G., & Choo, H. (2023). *GSCA Pro for Windows User's Manual*. https://www.gscapro.com/_files/ugd/7b7a8b_5069efdf6afd4686b9463964a9352796.pdf

Kirova, A. M., Bays, R. B., & Lagalwar, S. (2015). Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. *BioMed Research International*, *2015*. https://doi.org/10.1155/2015/748212

Málišová, E., Dančík, D., Heretik, A., Abrahámová, M., Krakovská, S., Brandoburová, P., & Hajdúk, M. (2021). Slovak version of the Trail Making Test: Normative data. *Applied Neuropsychology: Adult*, *29*(6), 1476–1483. https://doi.org/10.1080/23279095.2021.1890596

Mendez, M. F. (2017). Early-onset Alzheimer disease. *Neurologic clinics*, *35*(2), 263–281. https://doi.org/10.1016/j.ncl.2017.01.005

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Ojeda, N., Aretouli, E., Peña, J., & Schretlen, D. J. (2014). Age differences in cognitive performance: A study of cultural differences in Historical Context. *Journal of Neuropsychology*, *10*(1), 104–115. https://doi.org/10.1111/jnp.12059

Osuka, Y., Kojima, N., Sakurai, R., Watanabe, Y., & Kim, H. (2020). Reliability and construct validity of a novel motor–cognitive dual-task test: A Stepping Trail Making Test. *Geriatrics and Gerontology International*, *20*(4), 291–296. https://doi.org/10.1111/ggi.13878

Price, L. R. (2017). *Psychometric methods: Theory into practice*. Guilford Publications, Inc. www.guilford.com/MSS

Prince, M. (2015). World Alzheimer report. https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf

Reitan, R. M., & Wolfson, D. (2004). The Trail Making Test as an initial screening procedure for neuropsychological impairment in older children. *Archives of Clinical Neuropsychology*, *19*(2), 281–288. https://doi.org/10.1016/S0887-6177(03)00042-8

Seo, E. H., Lee, D. Y., Kim, K. W., Lee, J. H., Jhoo, J. H., Youn, J. C., Choo, I. H., Ha, J., & Woo, J. I. (2006). A normative study of the trail making test in Korean elders. *International Journal of Geriatric Psychiatry*, *21*(9), 844–852. https://doi.org/10.1002/gps.1570

Siciliano, M., Chiorri, C., Battini, V., Sant'Elia, V., Altieri, M., Trojano, L., & Santangelo, G. (2019). Regression-based normative data and equivalent scores for Trail Making Test (TMT): An updated Italian normative study. *Neurological Sciences*, *40*(3), 469–477. https://doi.org/10.1007/s10072-018-3673-y

Souza, A. C. d., Alexandre, N. M. C., Guirardello, E. d. B., Souza, A. C. d., Alexandre, N. M. C., & Guirardello, E. d. B. (2017). Propriedades psicométricas na avaliaccão de instrumentos: Avaliaccão da confiabilidade e da validade. *Epidemiologia e Serviccos de Saúde*, *26*(3), 649–659. https://doi.org/10.5123/s1679-49742017000300022

Specka, M., Weimar, C., Stang, A., Jockel, K. H., Scherbaum, N., Hoffmann, S. S., Kowall, B., & Jokisch, M. (2022). Trail making test normative data for the German older population. *Archives of Clinical Neuropsychology*, *37*(1), 186–198. https://doi.org/10.1093/arclin/acab027

Suriastini, Turana, Y., Witoelar, F., Supraptilah, B., Wicaksono, T. Y., & Dwi, E. (2018). Angka prevalensi demensia, Perlu perhatian kita semua [Dementia prevalence rate, We all need to pay attention]. *Pancanaka*, *1*(2), 14.

Suwartono, C., Halim, M. S., Hidajat, L. L., Hendriks, M. P. H., & Kessels, R. P. C. (2014). Development and reliability of the Indonesian Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV). *Psychology*, *05*(14), 1611–1619. https://doi.org/10.4236/psych.2014.514171

Sweetasari, A. G. (2022). Sosialisasi dan pelatihan pemeriksaan fungsi kognitif untuk deteksi dini penyakit demensia pada tenaga kesehatan di Puskesmas wilayah Dinkes kota Cimahi [Socialization and training of cognitive function examination for early detection of dementia in health workers at the Community Health Center in the Cimahi City Health Office area]. *Jurnal Abdimas Kartika Wijayakusuma*, *3*(2), 178–186. https://doi.org/10.26874/jakw.v3i2.261

Trull, T. J., & Prinstein, M. J. (2013). *Clinical psychology*. Jon-David Hague.

Turana, Y., Tengkawan, J., Suswanti, I., Suharya, D., Riyadina, W., & Pradono, J. (2019). Primary prevention of Alzheimer's disease in Indonesia. *International Journal of Aging Research*. https://doi.org/10.28933/ijoar-2019-06-2506

Wagner, S., Helmreich, I., Dahmen, N., Lieb, K., & Tadi, A. (2011). Reliability of three alternate forms of the trail making tests A and B. *Archives of Clinical Neuropsychology*, *26*(4), 314–321. https://doi.org/10.1093/arclin/acr024

Wahyuningrum, S. E., Sulastri, A., Hendriks, M., Consortium, I. N., & van Luijtelaar, G. (2022). The Indonesian Neuropsychological Test Battery (INTB): The underlying cognitive constructs, and the effects. *Acta Neuropsychologica*, *20*(4), 445–470. https://doi.org/10.5604/01.3001.0016.1339

Widhianingtanti, L. T., Luijtelaar, G. V., Suryani, A. O., Hestyanti, Y. R., & Sulastri, A. (2022). Indonesian trail making test: Analysis of psychometric properties, effects of demographic variables, and norms for Javanese adults. *Jurnal Psikologi*, *49*(2), 104. https://doi.org/10.22146/jpsi.68953

Wulanyani, N. M. S., Widhianingtanti, L. T., Immanuel, A. S., Aisyah, A. R. K., Hendriks, M. P. H., Hestyanti, Y. R., Suryani, A. O., & Van Luijtelaar, G. (2024). Psychometric properties of the Five-executive Function Tests in Indonesian samples. *Psikohumaniora: Jurnal Penelitian Psikologi*, *9*(1), 125–146. https://doi.org/10.21580/pjpp.v9i1.20957

Zucchella, C., Federico, A., Martini, A., Tinazzi, M., Bartolo, M., & Tamburin, S. (2018). Neuropsychological testing. *Practical Neurology*, *18*(3), 227–237. https://doi.org/10.1136/practneurol-2017-001743